

Large Deviation Results for the Nonparametric Regression Function Estimator on Functional Data

D. LOUANI^{a*} & S. M. OULD MAOULOUD^b

^a Université de Reims and L.S.T.A., Université de Paris 6, France.

^b École des Mines de Mauritanie, Mauritania.

Abstract - This paper is devoted to the study of large deviation behaviors in the setting of the estimation of the regression function on functional data. A large deviation principle is stated for a process Z_n , defined below, allowing to derive a pointwise large deviation principle for the Nadaraya-Watson-type l -indexed regression function estimator as a by-product. Moreover, a uniform over VC-classes Chernoff type large deviation result is stated for the deviation of the l -indexed regression estimator.

Key words: Functional data, entropy, kernel estimator, large deviation, regression function, vc-classes.

2000 Mathematics Subject Classifications: 60F10, 62G07, 62F05, 62H15.

1. INTRODUCTION

The regression problem has received a great interest and has motivated a great number of investigations and studies throughout the time bringing to the statistic literature a considerable knowledge. A number of models and nonparametric estimators to estimate the regression function have been proposed in the literature when the discrete time or the continuous time explanatory random variables take their values in a finite dimensional space where the Lebesgue measure plays an important role. We refer to Bosq (1998) for an account of properties and results and the

*Corresponding author email : djamal.louani@upmc.fr

references therein. Due to the availability of computing resources that allow sharp recordings in phenomena observation up to the level where data may be treated as curves, functional modeling has received a lot of attention in the last few years from mathematical, probabilistic, statistical or physicist points of view. It is worth noticing that there is an increasing number of sources of potential applications of functional models, as in chemometrics, environmetrics, speech recognition, radar range profile studies, medical data and so on. The number of publications studying properties of these models, as asymptotic issues for example, grows continuously. For an overview of the present state of the art, we refer to the works of Gasser *et al* (1998), Bosq (2000), Ferraty and Vieu (2000, 2004), Ramsay and Silverman (2002, 2005), Masry (2005), Ferraty *et al* (2007), Ezzahrioui *et al* (2008), and to the recent monograph by Ferraty and Vieu (2006) and the references therein.

To introduce the study framework, let $(X_i, Y_i)_{i \in \mathbb{N}}$ be a sequence of i.i.d. pairs of random elements where Y_i is a real-valued random variable with density g , with respect to the Lebesgue measure on \mathbb{R} , and X_i takes its values in some semi-metric abstract space $(\mathcal{E}, d(\cdot, \cdot))$. This covers the case of semi-normed spaces of possibly infinite dimension (e.g., Hilbert or Banach spaces) with the norm $\|\cdot\|$ and the distance $d(x, y) = \|x - y\|$. For a real function l and any fixed $x \in \mathcal{E}$, the l -indexed regression function at $X_1 = x$ is defined by $r^l(x) := \mathbb{E}(l(Y_1)|X_1 = x)$.

The Nadaraya-Watson type estimator of r^l has been introduced by Ferraty and Vieu (2000). It is defined, for any fixed $x \in \mathcal{E}$, by

$$\hat{r}_n^l(x) = \begin{cases} \frac{\sum_{i=1}^n l(Y_i) K\left(\frac{d(x, X_i)}{h}\right)}{\sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right)} := \frac{\hat{r}_{n,2}(x)}{\hat{r}_{n,1}(x)}, & \text{if } \hat{r}_{n,1}(x) \neq 0, \\ 0, & \text{elsewhere.} \end{cases} \quad (1)$$

Here, K is a real-valued kernel function, $h := h_n$ is the bandwidth parameter (which goes to 0 as n goes to infinity),

$$\hat{r}_{n,1}(x) = \frac{1}{n\phi(h)} \sum_{i=1}^n \Delta_i(x) \quad \text{and} \quad \hat{r}_{n,2}(x) = \frac{1}{n\phi(h)} \sum_{i=1}^n l(Y_i) \Delta_i(x), \quad (2)$$

where

$$\Delta_i(x) = K\left(\frac{d(x, X_i)}{h}\right)$$

and ϕ is a positive function that will be defined below. Notice that the index function l allows to study simultaneously properties of several estimates. The first example is given by the most classical regression function estimator where l stands as the identity function. Whenever $l = \mathbb{1}_A$ is the indicator function of the set A , $\hat{r}_n^l(x)$ is

the estimator of the conditional probability measure of the event $\{Y_1 \in A\}$ given $X_1 = x$.

For $x \in \mathcal{E}$, consider now the vector process

$$Z_n(x) = (\hat{r}_{n,1}(x), \hat{r}_{n,2}(x)).$$

In this paper, we aim at establishing a large deviation principle for the process $Z_n(x)$ and deriving asymptotics for the l -indexed regression function estimator in both the pointwise case and the uniform, over some vc-classes, case.

There exists an extensive large deviation literature involving many areas of probability and statistics. We refer to the books of Dembo & Zeitouni (1998) and Deuschel & Stroock (1989) and the references therein for an account of results and applications. In nonparametric function estimation setting, several results have been obtained these last years. We refer to Louani (1999) and Ould Maouloud (2008) where the studies involve the Nadaraya-Watson and histogram estimates of the regression function respectively both in the real vector case. Using the delta-sequence estimation method, Louani and Ould Maouloud (2011) established a large deviation principle for the real regression function estimate embedded in the L_1 space equipped with the weak topology. Notice that the main applications of large deviation results are related to the efficiency of tests in the Bahadur sense, see Nikitin (1995) for more details, together with the inaccuracy rate of estimators that allow to compare testing procedures and estimation performances respectively. The results may be also used to establish estimates consistency with rates of convergence.

2. RESULTS

Our results are stated under some assumptions we gather hereafter for easy reference

- (A1) K is a nonnegative bounded differentiable kernel over its support $[0, 1]$ and $K(1) > 0$. The derivative K' of K exists on the interval $[0, 1]$.
- (A2) For $x \in \mathcal{E}$ and a real number v , there exist a nonnegative functional f_v and a nonnegative real function ϕ tending to zero, as its argument tends to 0, such that, uniformly in v ,
 - (i) $F_x(u|Y = v) = \mathbb{P}(d(x, X_1) \leq u|Y = v) = \phi(u)f_v(x) + o(\phi(u))$ as $u \rightarrow 0$,
 - (ii) There exists a nondecreasing bounded function τ_0 such that, uniformly in $u \in [0, 1]$,

$$\frac{\phi(hu)}{\phi(h)} = \tau_0(u) + o(1), \quad \text{as } h \downarrow 0,$$

(A3) For any real numbers a and b ,

$$(i) \int (e^{a+bl(v)} - 1) f_v(x) g(v) dv < \infty, \quad (ii) \int e^{al(v)} f_v(x) g(v) dv < \infty,$$

$$(iii) \int e^{al(v)} f_v(x) g(v) l(v) dv < \infty, \quad (iv) \int e^{al(v)} f_v(x) g(v) l^2(v) dv < \infty.$$

Discussions of hypotheses. Condition (A1) is very usual in nonparametric estimation literature devoted to functional data context. From the fact that Lebesgue measure does not exist on infinite dimension space, hypotheses (A2) involve the small ball techniques related to the fractal dimension used in this paper. A number of examples of the function ϕ together with the corresponding decomposition of the probability of the small balls are given throughout several works (See, e.g., Ferraty and Vieu (2000, 2004, 2006), Ferraty *et al* (2006), Ezzahrioui *et al* (2008) and Laïb and Louani (2010)). A further example is given hereafter to illustrate the condition (A2)(i). Hypotheses (A3) are set on to insure the needed properties of finiteness and differentiability of the moment generating function of the process $Z_n(x)$. These hypotheses induce the fact that the large deviation principle holds with a good rate function, a property that is strongly expected in such results.

Example 1. Let \mathcal{E} be a separate abstract space equipped with the semi-metric defined, for $(x, y) \in \mathcal{E}^2$, by

$$d(x, y) = \left| \int (x(t) - y(t)) dt \right|.$$

Consider two elements h and l in \mathcal{C} together with the regression model

$$X_i = Y_i h + \varepsilon_i l,$$

where ε_i is a real random variable independent of Y_i . Observe now, for any $u > 0$, that we have

$$F_x(u|Y = v) = \mathbb{P}(d(x, X_i) \leq u|Y_i = v) = \mathbb{P}\left(\left| \int (x(t) - X_i(t)) dt \right| \leq u|Y_i = v\right).$$

Consequently, while assuming $0 \neq \left| \int l(t) dt \right| < \infty$, $\left| \int x(t) dt \right| < \infty$ and $\left| \int h(t) dt \right| < \infty$, it follows that

$$F_x(u|Y = v) = \mathbb{P}\left(\left| \int (x(t) - Y_i h(t) - \varepsilon_i l(t)) dt \right| \leq u|Y_i = v\right)$$

$$\begin{aligned}
&= \mathbb{P} \left(\left| \int (x(t) - v h(t) - \varepsilon_i l(t)) dt \right| \leq u \right) \\
&= \mathbb{P} \left(\frac{-u + \int x(t) dt - v \int h(t) dt}{\int l(t) dt} \leq \varepsilon_i \leq \frac{u + \int x(t) dt - v \int h(t) dt}{\int l(t) dt} \right) \\
&= \Phi \left(\frac{u + \int x(t) dt - v \int h(t) dt}{\int l(t) dt} \right) - \Phi \left(\frac{-u + \int x(t) dt - v \int h(t) dt}{\int l(t) dt} \right),
\end{aligned}$$

where Φ is the distribution function of ε_i . Taking Φ as the $\mathcal{N}(0, 1)$ distribution function and assuming that $0 < \int l(t) dt < \infty$, we obtain,

$$F_x(u|Y = v) = \frac{u}{\int l(t) dt} \sqrt{\frac{2}{\pi}} \exp \left(-\frac{1}{2} \left(\frac{\int x(t) dt - v \int h(t) dt}{\int l(t) dt} \right)^2 \right) (1 + o(1)),$$

and the condition (A2)(i) is satisfied with $\phi(u) = 2u$ and

$$f_v(x) = \frac{1}{\int l(t) dt \sqrt{2\pi}} \exp \left(-\frac{1}{2} \left(\frac{\int x(t) dt - v \int h(t) dt}{\int l(t) dt} \right)^2 \right).$$

□

The first result states a LDP for the process $Z_n(x)$.

Theorem 2.1 *Under assumptions (A1)-(A3), $Z_n(x)$ satisfies a LDP with the speed $n\phi(h)$ and a good rate function given by*

$$\Gamma_x(\lambda_1, \lambda_2) = \sup_{t_1, t_2} \{ \lambda_1 t_1 + \lambda_2 t_2 - \Phi^x(t_1, t_2) \},$$

where

$$\begin{aligned}
\Phi^x(t_1, t_2) &= \int f_v(x) \left[(e^{(t_1 + t_2 l(v))K(1)} - 1) \right. \\
&\quad \left. - \int_0^1 (t_1 + t_2 l(v)) K'(u) e^{(t_1 + t_2 l(v))K(u)} \tau_0(u) du \right] g(v) dv.
\end{aligned}$$

Remark 2.1 *If we suppose that the function τ_0 is differentiable, then, integrating by parts, we obtain*

$$\Phi^x(t_1, t_2) = \int \int_0^1 f_v(x) \tau_0'(u) (e^{(t_1 + t_2 l(v))K(u)} - 1) g(v) du dv$$

which gives a more simpler form of the rate function.

Whenever we take the function K as the uniform kernel, we obtain a more explicit rate function. In order to display it, we introduce first some further notations. For any $x \in \mathcal{E}$ and any $t \in \mathbb{R}$, set

$$V_x(t) = \frac{\int e^{tl(u)} f_u(x) g(u) l(u) du}{\int e^{tl(v)} f_v(x) g(v) dv} \quad (3)$$

and

$$V_x^{-1}(t) = \inf\{s : V_x(s) \geq t\}.$$

Moreover, assuming that the derivative of the function τ_0 exists and considering the fact that $\tau_0(0) = 0$ and $\tau_0(1) = 1$, we observe, whenever $K(u) = \mathbb{1}_{[0,1]}(u)$, that

$$\Phi^x(t_1, t_2) = \int (e^{(t_1+t_2 l(v))} - 1) f_v(x) g(v) dv. \quad (4)$$

Corollary 2.1 *Assume that the function τ_0 is differentiable and that K is the uniform kernel over the interval $[0, 1]$, then, under assumptions (A3)(i)-(iv), we obtain the following explicit form of the rate function*

$$\Gamma_x(\lambda_1, \lambda_2) = \begin{cases} \lambda_1(\log \lambda_1 - 1) + \lambda_2 V_x^{-1}\left(\frac{\lambda_2}{\lambda_1}\right) - \lambda_1 \log \int e^{V_x^{-1}\left(\frac{\lambda_2}{\lambda_1}\right) l(v)} f_v(x) g(v) dv \\ + \int f_u(x) g(u) du & \text{if } \lambda_1 > 0 \text{ and } v_0(x) < \lambda_2/\lambda_1 < v_1(x) \\ \infty & \text{elsewhere,} \end{cases} \quad (5)$$

where $v_0(x) = \inf_{t \in \mathbb{R}} V_x(t)$ and $v_1(x) = \sup_{t \in \mathbb{R}} V_x(t)$.

Remark 2.2 *Whenever $l = \mathbb{1}_A$ is the indicator function of a subset A of \mathbb{R} , it is possible to display a more explicit rate function whenever τ_0 is differentiable. Towards this end, for any $B \subset \mathbb{R}$ and any $t \in \mathbb{R}$, set*

$$W_x(B) = \int_B f_v(x) g(v) dv \quad \text{and} \quad \zeta(t) = \int_0^1 \tau'_0(u) K(u) e^{tK(u)} du.$$

It follows then that

$$\begin{aligned} \Gamma_x(\lambda_1, \lambda_2) &= (\lambda_1 - \lambda_2) \zeta^{-1} \left(\frac{\lambda_1 - \lambda_2}{W_x(\bar{A})} \right) + \lambda_2 \zeta^{-1} \left(\frac{\lambda_2}{W_x(A)} \right) + \int f_v(x) g(v) dv \\ &\quad - \int_0^1 \tau'_0(u) \left[W_x(A) \exp \left\{ \zeta^{-1} \left(\frac{\lambda_2}{W_x(A)} \right) K(u) \right\} \right. \\ &\quad \left. + W_x(\bar{A}) \exp \left\{ \zeta^{-1} \left(\frac{\lambda_1 - \lambda_2}{W_x(\bar{A})} \right) K(u) \right\} \right] du, \end{aligned}$$

where \bar{A} is the complementary set of A and

$$\zeta^{-1}(t) = \inf\{s : \zeta(s) \geq t\}.$$

The following corollary gives the result pertaining to a large deviation principle for the regression function estimate at the point x .

Corollary 2.2 *Under hypotheses of Theorem 2.1, the regression function estimate $\hat{r}_n^l(x)$ satisfies a LDP with the speed $n\phi(h)$ and the good rate function defined by*

$$\gamma_x(\lambda) = \inf_{\lambda_1} \{\Gamma_x(\lambda_1, \lambda \times \lambda_1)\}.$$

Remark 2.3 *If we assume that the function τ_0 is differentiable and that K is the uniform kernel over the interval $[0, 1]$, then we obtain the following explicit form of the rate function*

$$\gamma_x(\lambda) = \int (1 - \exp\{V_x^{-1}(\lambda)(l(v) - \lambda)\}) f_v(x)g(v)dv, \quad (6)$$

whenever $v_0(x) < \lambda < v_1(x)$, and $\gamma_x(\lambda) = \infty$ elsewhere.

Remark 2.4 *The first and second derivatives of the function γ_x given in the statement (6) are*

$$\gamma'_x(\lambda) = V_x^{-1}(\lambda) \exp\{-\lambda V_x^{-1}(\lambda)\} \int e^{V_x^{-1}(\lambda)l(v)} f_v(x)g(v)dv$$

and

$$\gamma''_x(\lambda) = \left(\frac{1}{V'_x(V_x^{-1}(\lambda))} - (V_x^{-1}(\lambda))^2 \right) \exp\{-\lambda V_x^{-1}(\lambda)\} \int e^{V_x^{-1}(\lambda)l(v)} f_v(x)g(v)dv$$

respectively. When Z denotes the random variable associated to the density function $f_v(x)g(v)/\int f_v(x)g(v)dv$, it follows that $V_x(0) = \mathbb{E}(l(Z))$. Therefore, assuming that $\mathbb{E}(l(Z)) = 0$, by Taylor series expansion we obtain, in the neighborhood of $\lambda = 0$, that

$$\gamma_x(\lambda) = \frac{\lambda^2}{2 \mathbb{E}(l^2(Z))} \int f_v(x)g(v)dv(1 + o(1)).$$

In the sequel, we investigate the uniform aspects of large deviation, in the Chernoff sense, of the regression function estimate $\hat{r}_n^l(x)$. More precisely, we consider the asymptotic behavior of the quantity $\|\hat{r}_n^l - r^l\|_{\mathcal{C}} := \sup_{x \in \mathcal{C}} |\hat{r}_n^l(x) - r^l(x)|$, where \mathcal{C} is a class of elements of \mathcal{E} . Towards this end, for any $\varepsilon > 0$, consider the following number

$$\mathcal{N}(\varepsilon, \mathcal{C}, d) = \min\{n : \text{there exist } c_1 \cdots, c_n \text{ in } \mathcal{C} \text{ such that } \forall x \in \mathcal{C} \\ \text{there exists } 1 \leq k \leq n \text{ such that } d(x, c_k) < \varepsilon\}$$

which measures how full is the class \mathcal{C} . Further notations are needed to display the uniform large deviation result. From now on, set

$$\beta(x, \lambda) = \inf\{\gamma_x(\alpha + r^l(x)) : \alpha \in (-\infty, -\lambda] \cup [\lambda, \infty)\}, \quad (7)$$

$\rho(\lambda) = \inf_{x \in \mathcal{C}} \beta(x, \lambda)$, and D_Y to be the set of values of the random variable Y_1 . Moreover consider the conditions

- (A4) (i) $\sup_{x \in \mathcal{C}} \int |l(v)| f_v(x) g(v) dv < \infty$,
(ii) The condition (A2)(i) is satisfied uniformly in $v \in D_Y$,
(iii) $\int_0^1 |K'(u)| \tau_0(u) du < \infty$.

Remark 2.5 *In the setting of Remark 2.3, it is easily seen that*

$$\beta(x, \lambda) = \min\{\gamma_x(-\lambda + r^l(x)), \gamma_x(\lambda + r^l(x))\},$$

since the function γ_x is non-increasing on the left of $r^l(x)$ and non-decreasing on its right.

The following theorem gives a Chernoff-type large deviation result for the uniform deviation of the estimate \hat{r}_n^l with respect to r^l .

Theorem 2.2 *Suppose that the function r^l is uniformly continuous upon \mathcal{C} and that the kernel K is a Lipschitz function bounded from below by a constant $K_0 > 0$. Under hypotheses (A1)-(A4) whenever the condition*

$$\nu = o\left(\frac{nh}{\exp\{An\phi(h)\}}\right), \text{ for any } A > 0, \text{ and } \lim_{n \rightarrow \infty} \frac{\log \mathcal{N}(\nu, \mathcal{C}, d)}{n\phi(h)} = 0 \quad (8)$$

is satisfied and the function ρ is continuous, for any $\lambda > 0$, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n\phi(h)} \log \mathbb{P}(\|\hat{r}_n^l - r^l\|_{\mathcal{C}} > \lambda) = -\rho(\lambda). \quad (9)$$

The continuity of the rate function ρ is a needed condition to obtain the result of Theorem 2.2. It is then natural to ask the question about the required assumptions for this condition to be satisfied. The following propositions give a reply to this question.

Proposition 2.1 *Assume that the parametric family of functions $\{\Gamma_x(\lambda, \mu)\}_{x \in \mathcal{C}}$ is equi-continuous and that the function $r^l(x)$ is bounded. Then, the function ρ is continuous.*

It is difficult to state the conditions under which the family of functions $\{\Gamma_x(\lambda, \mu)\}_{x \in \mathcal{C}}$ is equi-continuous in the general framework. Hereafter, we limit ourselves to the case where the kernel K is uniform over the interval $[0, 1]$ and the function τ_0 is differentiable as in the setting of Corollary 2.1. Towards this end, we first introduce the following notations. From now on, \bar{S} stands as the generic notation of the complementary of any set S , $d(v) = \inf_{x \in \mathcal{C}} f_v(x)$, $D(v) = \sup_{x \in \mathcal{C}} f_v(x)$, $A_l = \{v : l(v) - 1 \leq 0\}$, $B_l = \{v : l(v) + 1 \leq 0\}$ and $C_l = \{v : l(v) \leq 0\}$. Furthermore, we consider these additional assumptions

- (A5) (i) Whenever S stands as one of the sets A_l or B_l , we have $0 < \int_S d(v)g(v)dv < \infty$ and $0 < \int_{\bar{S}} d(v)g(v)dv < \infty$,
- (ii) For any t , $\left| \int e^{tl(v)} l(v) (D(v)\mathbf{1}_{C_l}(v) + d(v)\mathbf{1}_{\bar{C}_l}(v)) g(v) dv \right| < \infty$,
- (iii) For any t , $\left| \int e^{tl(v)} l(v) (d(v)\mathbf{1}_{C_l}(v) + D(v)\mathbf{1}_{\bar{C}_l}(v)) g(v) dv \right| < \infty$,
- (iv) $\int_{C_l} e^{tl(v)} (\mathbf{1}_{\{t \leq 0\}} \mathbf{1}_{C_l}(v) + \mathbf{1}_{\{t > 0\}} \mathbf{1}_{\bar{C}_l}(v)) D(v) g(v) dv < \infty$.

Proposition 2.2 *In the setting of Corollary 2.1, assume that assumptions (A5) are satisfied. Then the family of functions $\{\Gamma_x(\lambda, \mu)\}_{x \in \mathcal{C}}$ is equi-continuous on its finiteness domain.*

Remark 2.6 *Whenever the smoothing parameter h and the function ϕ are such that*

$$\frac{\exp\{An\phi(h)\}}{n\phi(h)} = o(nh), \quad \text{for any } A > 0, \quad (10)$$

then the condition (8) takes the form

$$\lim_{\nu \rightarrow 0} \nu \log \mathcal{N}(\nu, \mathcal{C}, d) = 0. \quad (11)$$

Notice that the condition (10) is satisfied when, for example, we take $\phi(h) = ah^\alpha$ and $h = \left(\frac{\log \log n}{n}\right)^{\frac{1}{\alpha}}$ with $\alpha > 1$ and $a > 0$. The condition (11) is very usual in defining Vapnik-Chervonenkis classes, see, for instance van der Vaart and Wellner (1996). Hereafter, examples of classes fulfilling the condition (11) are displayed.

EXAMPLE 1. (*Parametric classes of functions*) For a function \mathcal{X} in the space $L_p(\mathbb{R})$, consider the parametric class of functions defined by

$$\mathcal{C} = \{\mathcal{X}_a(\cdot) = a\mathcal{X}(a\cdot) : -\infty < A_1 \leq a \leq A_2 < \infty \text{ and } a \neq 0\}$$

together with the L_p -distance given, for $1 \leq p < \infty$, by

$$d_p(x, y) = \left(\int |x(t) - y(t)|^p dt \right)^{\frac{1}{p}}.$$

assuming that the function \mathcal{X} is differentiable with a continuous derivative and that the function $q(t) = t\mathcal{X}'(t)$ is L_p -integrable, it follows that there exists a positive constant C_0 such that $N(\nu, \mathcal{C}, d_p) \leq C_0/\nu$. Therefore, the condition (11) is satisfied.

EXAMPLE 2. (*Classes of functions that are Lipschitz in a parameter*) Let T be an index set and consider d a distance over T . Suppose that $\mathcal{C} = \{x_t : t \in T\}$ is a class of functions defined on \mathbb{R} , that are Lipschitz in the index parameter t in the sense that there exists a function x on \mathbb{R} such that for any $u \in \mathbb{R}$,

$$|x_s(u) - x_t(u)| \leq d(s, t)x(u).$$

From Theorem 2.7.11 in van der Vaart & Wellner (1996), it follows, for any norm $\|\cdot\|$ whenever $\|x\| < \infty$, that

$$\mathcal{N}(\nu\|x\|, \mathcal{C}, \|\cdot\|) \leq N(\nu, T, d),$$

where $N(\nu, T, d)$ is the minimal number of balls of radius ν needed to cover T . Therefore, the condition (11) may be expressed as

$$\lim_{\nu \rightarrow 0} \nu \log N(\nu, T, d) = 0,$$

which is a Vapnik-Chervonenkis class of sets condition. Naturally, it is more easy to display examples of Vapnik-Chervonenkis classes of sets. As an example, let x be a Lipschitz boundedly supported function. For an index set T included in \mathbb{R} , define the class \mathcal{C} by taking, for any $u \in \mathbb{R}$, $x_t(u) = x(u - t)$. It is obvious then that the x_t 's are Lipschitz with respect to the index parameter t . Taking d as the absolute distance on T that we take as a bounded convex interval, it follows easily that $N(\nu, T, d) \leq |T|/\nu$, where $|T|$ is the diameter of T . Therefore, the condition (11) is satisfied.

EXAMPLE 3. (*Smooth function classes*) Let \mathcal{U} be a bounded convex subset of \mathbb{R}^d with nonempty interior. For any $\alpha > 0$, consider the class of functions on \mathcal{U} that possess uniformly bounded derivatives up to order $[\alpha]$, where $[\alpha]$ stands as the integer part of α , and whose highest derivatives are Lipschitz of order $\alpha - [\alpha]$. Denote by $x^{(l)}$ the l -th derivative of x and set, for any $x : \mathcal{U} \rightarrow \mathbb{R}$,

$$\|x\|_\alpha := \max_{l \leq [\alpha]} \sup_{u \in \mathcal{U}} |x^{(l)}(u)| + \sup_{u, v \in \overset{o}{\mathcal{U}}, u \neq v} \frac{|x^{([\alpha])}(u) - x^{([\alpha])}(v)|}{|u - v|^{\alpha - [\alpha]}}$$

where $\overset{o}{\mathcal{U}}$ is the interior of \mathcal{U} . Let $\mathcal{C} = C_M^\alpha(\mathcal{U})$ be the class of all continuous functions $x : \mathcal{U} \rightarrow \mathbb{R}$ with $\|x\|_\alpha \leq M$. It follows from Corollary 2.7.2 of van der Vaart & Wellner (1996) that, for every $p \geq 1$ and any $\nu > 0$, there exists a constant c depending only on α , the diameter of \mathcal{U} and d such that

$$\log N_{[\cdot]}(\nu, C_1^\alpha(\mathcal{U}), L_p) \leq c \left(\frac{1}{\nu} \right)^{\frac{d}{\alpha}}.$$

Here, $\mathcal{N}_{[\cdot]}$ denotes the bracketing number (see, for instance, van der Vaart & Wellner (1996) page 83, for the definition) and, for a probability measure μ , $L_p(x, y) = (\int |x(u) - y(u)|^p d\mu(u))^{\frac{1}{p}}$. Since

$$N(\nu, C_1^\alpha(\mathcal{U}), L_p) \leq N_{[\cdot]}(2\nu, C_1^\alpha(\mathcal{U}), L_p),$$

it is then clear that the condition (11) is satisfied provided that $\alpha > 1$.

3. PROOFS

Proof of Theorem 2.1 The Laplace transform associated to the process $n\phi(h)Z_n(x)$ is defined, for any $(t_1, t_2) \in \mathbb{R}^2$, by

$$\Phi_n^x(t_1, t_2) = \mathbb{E}[\exp\{\langle (t_1, t_2), n\phi(h)(\hat{r}_{n,1}(x), \hat{r}_{n,2}(x)) \rangle\}]$$

$$\begin{aligned}
&= \mathbb{E} \left[\exp \left\{ \left\langle (t_1, t_2), \left(\sum_{i=1}^n \Delta_i(x), \sum_{i=1}^n l(Y_i) \Delta_i(x) \right) \right\rangle \right\} \right] \\
&= \mathbb{E} \left[\exp \left\{ \sum_{i=1}^n (t_1 + t_2 l(Y_i)) \Delta_i(x) \right\} \right] \\
&= (\mathbb{E} [\exp \{ (t_1 + t_2 l(Y_i)) \Delta_1(x) \}])^n := (\varphi_n^x(t_1, t_2))^n,
\end{aligned}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product. Let us now evaluate the quantity $\varphi_n^x(t_1, t_2)$. Observe that

$$\begin{aligned}
\varphi_n^x(t_1, t_2) &= 1 + \mathbb{E} [\exp \{ (t_1 + t_2 l(Y_1)) \Delta_1(x) \} - 1] \\
&= 1 + \int_0^1 \int (e^{(t_1 + t_2 l(v))K(u)} - 1) d\mathbb{P} \left(\frac{d(x, X_1)}{h} \leq u, Y \leq v \right) \\
&= 1 + \int_0^1 \int (e^{(t_1 + t_2 l(v))K(u)} - 1) d\mathbb{P} \left(\frac{d(x, X_1)}{h} \leq u \mid Y = v \right) d\mathbb{P}(Y \leq v)
\end{aligned}$$

Integrating by parts with respect to the component u , we obtain

$$\begin{aligned}
\varphi_n^x(t_1, t_2) &= 1 + \int [(e^{(t_1 + t_2 l(v))K(1)} - 1) F_x(h \mid Y = v) \\
&\quad - \int_0^1 (t_1 + t_2 l(v)) K'(u) e^{(t_1 + t_2 l(v))K(u)} F_x(uh \mid Y = v) du] d\mathbb{P}(Y \leq v).
\end{aligned}$$

Making use of the condition (A2)(i), we obtain

$$\begin{aligned}
\varphi_n^x(t_1, t_2) &= 1 + \int [(e^{(t_1 + t_2 l(v))K(1)} - 1) (\phi(h) f_v(x) + o(\phi(h))) \\
&\quad - \int_0^1 (t_1 + t_2 l(v)) K'(u) e^{(t_1 + t_2 l(v))K(u)} (\phi(uh) f_v(x) + o(\phi(uh))) du] d\mathbb{P}(Y \leq v) \\
&= 1 + \phi(h) \left(\int [(e^{(t_1 + t_2 l(v))K(1)} - 1) (f_v(x) + o(1)) \right. \\
&\quad \left. - \int_0^1 (t_1 + t_2 v) K'(u) e^{(t_1 + t_2 l(v))K(u)} (f_v(x) + o(1)) \frac{\phi(uh)}{\phi(h)} du] d\mathbb{P}(Y \leq v) \right).
\end{aligned}$$

By the condition (A2)(ii), it follows that

$$\begin{aligned}
\varphi_n^x(t_1, t_2) &= 1 + \phi(h) \left(\int [(e^{(t_1 + t_2 l(v))K(1)} - 1) (f_v(x) + o(1)) \right. \\
&\quad \left. - \int_0^1 (t_1 + t_2 l(v)) K'(u) e^{(t_1 + t_2 l(v))K(u)} (f_v(x) + o(1)) (\tau_0(u) + o(1)) du] d\mathbb{P}(Y \leq v) \right).
\end{aligned}$$

Therefore, after a Taylor series expansion of the function $\log(1+u)$ around $u=0$, we obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n\phi(h)} \log \Phi_n^x(t_1, t_2) &:= \Phi^x(t_1, t_2) = \int f_v(x) \left[(e^{(t_1+t_2l(v))K(1)} - 1) \right. \\ &\quad \left. - \int_0^1 (t_1 + t_2l(v))K'(u)e^{(t_1+t_2l(v))K(u)}\tau_0(u)du \right] g(v)dv \end{aligned}$$

Note that the condition (A3) implies that the function $\Phi^x(t_1, t_2)$ is finite and differentiable everywhere. The Fenchel-Legendre transform of $\Phi^x(t_1, t_2)$ is given by

$$\Gamma_x(\lambda_1, \lambda_2) = \sup_{t_1, t_2} \{ \lambda_1 t_1 + \lambda_2 t_2 - \Phi^x(t_1, t_2) \}.$$

We have now to establish that the function $\Phi^x(t_1, t_2)$ is essentially smooth and to use the Gärtner-Ellis Theorem (see, Dembo & Zeitouni (1998), page 44) to achieve the proof.

Considering hypotheses (A3) (i)-(ii), it is clear that the interior of the set $D = \{(t_1, t_2) : \Phi^x(t_1, t_2) < \infty\}$ is not empty. Moreover, making use of conditions (A3), it follows that the function $\Phi^x(t_1, t_2)$ is differentiable throughout the domain $\overset{\circ}{D}$. Subsequently, it is clear that the function $\Phi^x(t_1, t_2)$ is steep and, therefore, is essentially smooth. \square

Proof of Corollary 2.1 In view of the statement (4), we have to maximize the function

$$Q(t_1, t_2) = \lambda_1 t_1 + \lambda_2 t_2 - \int (e^{(t_1+t_2l(v))} - 1) f_v(x)g(v)dv.$$

Since the function Q is concave, it is easily seen that its maximum is reached at the point

$$(t_1, t_2) = \left(\log(\lambda_1) - \log \int e^{V_x^{-1}\left(\frac{\lambda_2}{\lambda_1}\right)l(v)} f_v(x)g(v)dv, V_x^{-1}\left(\frac{\lambda_2}{\lambda_1}\right) \right)$$

which gives the main form of the rate function given in the statement (5).

In order to display the finiteness domain of the function Γ_x , we have to study the function V_x . Observe from hypotheses (A3) that V_x is a differentiable function and that its derivative is given by

$$\begin{aligned} V_x'(t) &= \frac{(\int e^{tl(v)} f_v(x)g(v)l^2(v)dv) (\int e^{tl(v)} f_v(x)g(v)dv) - (\int e^{tl(v)} f_v(x)g(v)l(v)dv)^2}{(\int e^{tl(v)} f_v(x)g(v)dv)^2} \\ &= \frac{\int \int e^{t(l(u)+l(v))} f_u(x)f_v(x)g(u)g(v)(l^2(v) - l(u)l(v))dudv}{(\int e^{tl(v)} f_v(x)g(v)dv)^2} \end{aligned}$$

$$= \frac{1}{2} \frac{\int \int e^{t(l(u)+l(v))} f_u(x) f_v(x) g(u) g(v) (l(v) - l(u))^2 du dv}{\left(\int e^{tl(v)} f_v(x) g(v) dv \right)^2} \geq 0.$$

Therefore, V_x is an increasing function. Notice that $v_1(x) = \lim_{t \rightarrow \infty} V_x(t) = \sup_t V_x(t)$ exists in the closure $\bar{\mathbb{R}}$ of \mathbb{R} .

Assuming now that $\lambda_2/\lambda_1 > v_1(x)$, it results that there exists $\varepsilon > 0$ such that, for any $t \in \mathbb{R}$,

$$\frac{\lambda_2}{\lambda_1} + \varepsilon \geq V_x(t). \quad (12)$$

Integrating in both sides of (12) with respect to t , it follows, for any $t \in \mathbb{R}$, that

$$\left(\frac{\lambda_2}{\lambda_1} + \varepsilon \right) t + c_0 \geq \log \int e^{tl(v)} f_v(x) g(v) dv,$$

with $c_0 = \log \int f_v(x) g(v) dv$. Therefore, for any $(t_1, t_2) \in \mathbb{R}^2$, we obtain

$$\begin{aligned} H(t_1, t_2) &:= \lambda_1 t_1 + \lambda_2 t_2 - \exp \left\{ t_1 + \left(\frac{\lambda_2}{\lambda_1} + \varepsilon \right) t + c_0 \right\} + e^{c_0} \\ &\leq \lambda_1 t_1 + \lambda_2 t_2 - \int (\exp \{t_1 + t_2 l(v)\} - 1) f_v(x) g(v) dv. \end{aligned}$$

Thus,

$$\sup_{t_1, t_2} H(t_1, t_2) \leq \Gamma_x(\lambda_1, \lambda_2).$$

Studying now the function H , it is easily seen that

$$\infty = \sup_{t_1, t_2} H(t_1, t_2) \leq \Gamma_x(\lambda_1, \lambda_2)$$

whenever $\lambda_2/\lambda_1 > v_1(x)$. Similarly, whenever $\lambda_2/\lambda_1 < v_0(x)$, we obtain $\infty = \Gamma_x(\lambda_1, \lambda_2)$. \square

Proof of Corollary 2.2 The proof follows straightforwardly from Theorem 2.1 by making use of the contraction principle with the following continuous function

$$\begin{aligned} H &: \mathbb{R}_+^* \times \mathbb{R} \rightarrow \mathbb{R} \\ (\lambda_1, \lambda_2) &\rightarrow \frac{\lambda_2}{\lambda_1}. \end{aligned} \quad (13)$$

Consequently, \hat{r}_n^l satisfies the LDP with the speed $n\phi(h)$ and the rate function

$$\begin{aligned} \gamma_x(\lambda) &:= \inf \{ \Gamma_x(\lambda_1, \lambda_2) : H(\lambda_1, \lambda_2) = \lambda \} = \inf \{ \Gamma_x(\lambda_1, \lambda_2) : \lambda_2/\lambda_1 = \lambda \} \\ &= \inf \{ \Gamma_x(\lambda_1, \lambda_1 \times \lambda) : \lambda_1 > 0 \}. \end{aligned}$$

□

Proof of Theorem 2.2 First of all, since the rate function γ_x is continuous, it follows by the contraction principle, used with the continuous map $y \rightarrow y - r^l(x)$, that for any $\lambda > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{n\phi(h)} \log \mathbb{P}(|\hat{r}_n^l(x) - r^l(x)| > \lambda) = -\beta(x, \lambda). \quad (14)$$

To state the uniform lower bound, it suffices to notice that for any $x \in \mathcal{C}$, we have

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n\phi(h)} \log \mathbb{P}(\|\hat{r}_n^l - r^l\| > \lambda) &\geq \liminf_{n \rightarrow \infty} \frac{1}{n\phi(h)} \log \mathbb{P}(|\hat{r}_n^l(x) - r^l(x)| > \lambda) \\ &\geq -\beta(x, \lambda). \end{aligned}$$

Therefore,

$$\liminf_{n \rightarrow \infty} \frac{1}{n\phi(h)} \log \mathbb{P}(\|\hat{r}_n^l - r^l\| > \lambda) \geq -\rho(\lambda). \quad (15)$$

Towards establishing the upper bound, observe first that

$$\begin{aligned} \|\hat{r}_n^l - r^l\|_{\mathcal{C}} &= \max_{1 \leq j \leq \mathcal{N}(\nu, \mathcal{C}, d)} \sup_{x \in B_d(c_j, \nu)} |\hat{r}_n^l(x) - r^l(x)|, \\ &\leq \max_{1 \leq j \leq \mathcal{N}(\nu, \mathcal{C}, d)} |\hat{r}_n^l(c_j) - r^l(c_j)| + \max_{1 \leq j \leq \mathcal{N}(\nu, \mathcal{C}, d)} \sup_{x \in B_d(c_j, \nu)} |\hat{r}_n^l(c_j) - \hat{r}_n^l(x)| \\ &\quad + \max_{1 \leq j \leq \mathcal{N}(\nu, \mathcal{C}, d)} \sup_{x \in B_d(c_j, \nu)} |r^l(c_j) - r^l(x)|, \end{aligned} \quad (16)$$

where $B_d(c_j, \nu) = \{x \in \mathcal{E} : d(c_j, x) \leq \nu\}$. Assuming that r^l is uniformly continuous on \mathcal{C} , it follows, for any $\varepsilon > 0$, that there exists $\nu > 0$ such that

$$\max_{1 \leq j \leq \mathcal{N}(\nu, \mathcal{C}, d)} \sup_{x \in B_d(c_j, \nu)} |r^l(c_j) - r^l(x)| < \varepsilon.$$

For any $x \in \mathcal{C}$ and any $c \in \mathcal{C}$, observe that

$$\begin{aligned} \hat{r}_n^l(x) - \hat{r}_n^l(c) &= \frac{\sum_{i=1}^n l(Y_i) \left(K\left(\frac{d(x, X_i)}{h}\right) - K\left(\frac{d(c, X_i)}{h}\right) \right)}{\sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right)} \\ &\quad + \sum_{i=1}^n l(Y_i) K\left(\frac{d(x, X_i)}{h}\right) \left[\frac{\sum_{i=1}^n \left(K\left(\frac{d(x, X_i)}{h}\right) - K\left(\frac{d(c, X_i)}{h}\right) \right)}{\sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right) \sum_{i=1}^n K\left(\frac{d(c, X_i)}{h}\right)} \right]. \end{aligned}$$

Assuming the kernel K to be a Lipschitz function, it follows that

$$\left| K\left(\frac{d(x, X_i)}{h}\right) - K\left(\frac{d(c, X_i)}{h}\right) \right| \leq \frac{M}{h} |d(x, X_i) - d(c, X_i)| \leq \frac{M}{h} d(x, c),$$

where M is a positive constant. Therefore, whenever the kernel K is bounded away from below by $K_0 > 0$, we have

$$\begin{aligned} \sup_{x \in B_d(c, \nu)} |\hat{r}_n^l(x) - \hat{r}_n^l(c)| &\leq \frac{M\nu}{K_0 nh} \sum_{i=1}^n |l(Y_i)| + \frac{M\nu}{K_0^2 nh} \sum_{i=1}^n |l(Y_i)| K\left(\frac{d(c, X_i)}{h}\right) \\ &\leq \frac{2M\nu}{K_0^2 nh} \sum_{i=1}^n |l(Y_i)| K\left(\frac{d(c, X_i)}{h}\right). \end{aligned}$$

Thus, for any $\varepsilon > 0$, by Markov's inequality, we obtain

$$\begin{aligned} \mathbb{P}\left(\sup_{x \in B_d(c, \nu)} |\hat{r}_n^l(x) - \hat{r}_n^l(c)| > \varepsilon\right) &\leq \mathbb{P}\left(\sum_{i=1}^n |l(Y_i)| K\left(\frac{d(c, X_i)}{h}\right) > \varepsilon \frac{K_0^2 nh}{2M\nu}\right) \\ &\leq \frac{2M\nu}{\varepsilon K_0^2 h} \mathbb{E}\left(|l(Y_1)| K\left(\frac{d(c, X_1)}{h}\right)\right) \end{aligned}$$

Proceeding now similarly as in the proof of Theorem 2.1, we obtain

$$\begin{aligned} \mathbb{E}\left(|l(Y_1)| K\left(\frac{d(c, X_1)}{h}\right)\right) &= \phi(h) \left(\int [|l(v)| K(1)(f_v(c) + o(1)) \right. \\ &\quad \left. - \int_0^1 |l(v)| K'(u)(f_v(c) + o(1))(\tau_0(u) + o(1)) du] g(v) dv \right) \\ &:= \phi(h) A, \end{aligned}$$

where, by conditions (A4), A is a finite constant. Consequently, there exists a positive constant C such that

$$\mathbb{P}\left(\max_{1 \leq j \leq \mathcal{N}(\nu, \mathcal{C}, d)} \sup_{x \in B_d(c_j, \nu)} |\hat{r}_n^l(x) - \hat{r}_n^l(c_j)| > \varepsilon\right) \leq C \mathcal{N}(\nu, \mathcal{C}, d) \phi(h) \frac{\nu}{h\varepsilon}.$$

Considering the decomposition in the statement (16), it is easily seen that

$$\mathbb{P}(\|\hat{r}_n^l - r^l\|_{\mathcal{C}} > \lambda) = \mathbb{P}\left(\max_{1 \leq j \leq \mathcal{N}(\nu, \mathcal{C}, d)} |\hat{r}_n^l(c_j) - r^l(c_j)| > \lambda - 2\varepsilon\right)$$

$$\begin{aligned}
& \times \left(1 + \frac{\mathbb{P} \left(\max_{1 \leq j \leq \mathcal{N}(\nu, \mathcal{C}, d)} \sup_{x \in B_d(c_j, \nu)} |\hat{r}_n^l(x) - \hat{r}_n^l(c_j)| > \varepsilon \right)}{\mathbb{P} \left(\max_{1 \leq j \leq \mathcal{N}(\nu, \mathcal{C}, d)} |\hat{r}_n^l(c_j) - r^l(c_j)| > \lambda - 2\varepsilon \right)} \right) \\
& \leq \mathbb{P} \left(\max_{1 \leq j \leq \mathcal{N}(\nu, \mathcal{C}, d)} |\hat{r}_n^l(c_j) - r^l(c_j)| > \lambda - 2\varepsilon \right) \\
& \times \left(1 + \frac{C\mathcal{N}(\nu, \mathcal{C}, d)\phi(h)\frac{\nu}{h\varepsilon}}{\mathbb{P} \left(\max_{1 \leq j \leq \mathcal{N}(\nu, \mathcal{C}, d)} |\hat{r}_n^l(c_j) - r^l(c_j)| > \lambda - 2\varepsilon \right)} \right)
\end{aligned}$$

Since, for any $u > 0$, $\log(1 + u) \leq u$, it is obvious then that

$$\begin{aligned}
\frac{1}{n\phi(h)} \log \mathbb{P} (\|\hat{r}_n^l - r^l\|_{\mathcal{C}} > \lambda) & \leq \frac{1}{n\phi(h)} \log \mathbb{P} \left(\max_{1 \leq j \leq \mathcal{N}(\nu, \mathcal{C}, d)} |\hat{r}_n^l(c_j) - r^l(c_j)| > \lambda - 2\varepsilon \right) \\
& + \frac{C\mathcal{N}(\nu, \mathcal{C}, d)\nu}{nh\varepsilon \mathbb{P} \left(\max_{1 \leq j \leq \mathcal{N}(\nu, \mathcal{C}, d)} |\hat{r}_n^l(c_j) - r^l(c_j)| > \lambda - 2\varepsilon \right)}
\end{aligned}$$

Therefore, from the statement (15), we obtain

$$\begin{aligned}
& \frac{1}{n\phi(h)} \log \mathbb{P} (\|\hat{r}_n^l - r^l\|_{\mathcal{C}} > \lambda) \\
& \leq \frac{1}{n\phi(h)} \log \mathbb{P} \left(\max_{1 \leq j \leq \mathcal{N}(\nu, \mathcal{C}, d)} |\hat{r}_n^l(c_j) - r^l(c_j)| > \lambda - 2\varepsilon \right) \\
& + \frac{C\mathcal{N}(\nu, \mathcal{C}, d)\nu}{nh\varepsilon} \exp \{n\phi(h)(\rho(\lambda) + o(1))\} \\
& \leq \frac{1}{n\phi(h)} \log \mathcal{N}(\nu, \mathcal{C}, d) \\
& + \sup_{x \in \mathcal{C}} \frac{1}{n\phi(h)} \log \mathbb{P} \left(\max_{1 \leq j \leq \mathcal{N}(\nu, \mathcal{C}, d)} |\hat{r}_n^l(x) - r^l(x)| > \lambda - 2\varepsilon \right) \\
& + \frac{C\mathcal{N}(\nu, \mathcal{C}, d)\nu}{nh\varepsilon} \exp \{n\phi(h)(\rho(\lambda) + o(1))\}.
\end{aligned}$$

Making use of the condition (8), it follows that

$$\limsup_{n \rightarrow \infty} \frac{1}{n\phi(h)} \log \mathbb{P} (\|\hat{r}_n^l - r^l\|_{\mathcal{C}} > \lambda) \leq -\rho(\lambda - 2\varepsilon).$$

The proof is achieved while making ε tend to zero since the function $\rho(\lambda)$ is continuous. \square

Proof of Proposition 2.1 Observe first, for any positive real numbers λ and λ_1 , that

$$\begin{aligned}\rho(\lambda) &= \inf_{x \in \mathcal{C}} \beta(x, \lambda) = \inf_{x \in \mathcal{C}} \{\beta(x, \lambda) - \beta(x, \lambda_1) + \beta(x, \lambda_1)\} \\ &\leq \inf_{x \in \mathcal{C}} \beta(x, \lambda_1) + \sup_{x \in \mathcal{C}} \{\beta(x, \lambda) - \beta(x, \lambda_1)\}.\end{aligned}$$

Therefore, we have

$$|\rho(\lambda) - \rho(\lambda_1)| \leq \sup_{x \in \mathcal{C}} |\beta(x, \lambda) - \beta(x, \lambda_1)|.$$

Considering the statement (7), we obtain for $\lambda < \lambda_1$

$$\begin{aligned}|\rho(\lambda) - \rho(\lambda_1)| &\leq \sup_{x \in \mathcal{C}} \left| \inf_{|\alpha| \geq \lambda} \gamma_x(\alpha + r^l(x)) - \inf_{|\alpha| \geq \lambda_1} \gamma_x(\alpha + r^l(x)) \right| \\ &\leq \sup_{x \in \mathcal{C}} \left| \inf_{|\alpha| \geq \lambda} (\gamma_x(\alpha + r^l(x)) - \gamma_x(\lambda + r^l(x))) \right. \\ &\quad \left. - \inf_{|\alpha| \geq \lambda_1} (\gamma_x(\alpha + r^l(x)) - \gamma_x(\lambda + r^l(x))) \right| \\ &\leq \sup_{x \in \mathcal{C}} \left\{ \inf_{|\alpha| \geq \lambda} |\gamma_x(\alpha + r^l(x)) - \gamma_x(\lambda + r^l(x))| \right. \\ &\quad \left. + \inf_{|\alpha| \geq \lambda_1} |\gamma_x(\alpha + r^l(x)) - \gamma_x(\lambda + r^l(x))| \right\} \\ &\leq 2 \sup_{x \in \mathcal{C}} |\gamma_x(\lambda + r^l(x)) - \gamma_x(\lambda_1 + r^l(x))|.\end{aligned}$$

Taking into account the shape of the rate function γ_x given in Corollary 2.2, we obtain

$$\begin{aligned}|\rho(\lambda) - \rho(\lambda_1)| &\leq 2 \sup_{x \in \mathcal{C}} \left| \inf_{\delta} \Gamma_x(\delta, \delta(\lambda + r^l(x))) - \inf_{\delta} \Gamma_x(\delta, \delta(\lambda_1 + r^l(x))) \right| \\ &\leq 2 \sup_{x \in \mathcal{C}} \left| \inf_{\delta} (\Gamma_x(\delta, \delta(\lambda + r^l(x))) - \Gamma_x(\lambda, \lambda(\lambda_1 + r^l(x)))) \right. \\ &\quad \left. - \inf_{\delta} (\Gamma_x(\delta, \delta(\lambda_1 + r^l(x))) - \Gamma_x(\lambda, \lambda(\lambda_1 + r^l(x)))) \right| \\ &\leq 2 \sup_{x \in \mathcal{C}} |\Gamma_x(\lambda, \lambda(\lambda + r^l(x))) - \Gamma_x(\lambda, \lambda(\lambda_1 + r^l(x)))| \\ &\quad + 2 \sup_{x \in \mathcal{C}} |\Gamma_x(\lambda_1, \lambda_1(\lambda_1 + r^l(x))) - \Gamma_x(\lambda, \lambda(\lambda_1 + r^l(x)))|.\end{aligned}$$

It suffices now to use the fact that $\{\Gamma_x(\lambda, \mu)\}_{x \in \mathcal{C}}$ is a equi-continuous family of functions and that the regression function r^l is bounded to achieve the proof. \square

Proof of Proposition 2.2 Observe by the condition (A3)(i)-(iii) that Φ^x is a differentiable function and that

$$\Gamma_x(\lambda, \mu) = \lambda s_x + \mu t_x - \Phi^x(s_x, t_x),$$

where s_x and t_x are solutions of equations

$$(i) \quad \frac{\partial \Phi^x}{\partial s}(s, t) = \lambda, \quad (ii) \quad \frac{\partial \Phi^x}{\partial t}(s, t) = \mu. \quad (17)$$

respectively. Therefore, for any λ_1 and μ_1 , we have

$$\begin{aligned} \Gamma_x(\lambda, \mu) &= \lambda_1 s_x + \mu_1 t_x - \Phi^x(s_x, t_x) + (\lambda - \lambda_1)s_x + (\mu - \mu_1)t_x \\ &\leq \sup_{s, t} \{ \lambda_1 s + \mu_1 t - \Phi^x(s, t) \} + |(\lambda - \lambda_1)s_x| + |(\mu - \mu_1)t_x| \\ &\leq \Gamma_x(\lambda_1, \mu_1) + |(\lambda - \lambda_1)s_x| + |(\mu - \mu_1)t_x| \end{aligned}$$

and then

$$|\Gamma_x(\lambda, \mu) - \Gamma_x(\lambda_1, \mu_1)| \leq |\lambda - \lambda_1||s_x| + |\mu - \mu_1||t_x|.$$

We have now to state that both $\sup_{x \in \mathcal{C}} |t_x|$ and $\sup_{x \in \mathcal{C}} |s_x|$ are finite.

Observe from the statement (17)(i), whenever $t_x \geq 0$, that

$$\begin{aligned} \lambda &= e^{s_x} \left(\int_{A_l} e^{t_x + t_x(l(v)-1)} f_v(x) g(v) dv + \int_{\bar{A}_l} e^{t_x + t_x(l(v)-1)} f_v(x) g(v) dv \right) \\ &\geq e^{s_x + t_x} \int_{\bar{A}_l} e^{t_x(l(v)-1)} f_v(x) g(v) dv \geq e^{s_x + t_x} \int_{\bar{A}_l} d(v) g(v) dv. \end{aligned}$$

Therefore, we have

$$s_x + t_x \leq \log \left(\frac{\lambda}{\int_{\bar{A}_l} d(v) g(v) dv} \right).$$

Whenever $t_x < 0$, similarly, we obtain

$$s_x + t_x \leq \log \left(\frac{\lambda}{\int_{A_l} d(v) g(v) dv} \right).$$

Therefore, whatever the value that may take t_x , we obtain

$$s_x + t_x \leq \max \left\{ \log \left(\frac{\lambda}{\int_{\bar{A}_l} d(v) g(v) dv} \right), \log \left(\frac{\lambda}{\int_{A_l} d(v) g(v) dv} \right) \right\}. \quad (18)$$

On another hand, whenever $t_x \geq 0$, we have

$$\begin{aligned}\lambda &= e^{s_x} \left(\int_{B_l} e^{-t_x+t_x(l(v)+1)} f_v(x)g(v)dv + \int_{\bar{B}_l} e^{-t_x+t_x(l(v)+1)} f_v(x)g(v)dv \right) \\ &\geq e^{s_x-t_x} \int_{\bar{B}_l} e^{t_x(l(v)+1)} f_v(x)g(v)dv \geq e^{s_x-t_x} \int_{\bar{B}_l} d(v)g(v)dv.\end{aligned}$$

Consequently, we have

$$s_x - t_x \leq \log \left(\frac{\lambda}{\int_{\bar{B}_l} d(v)g(v)dv} \right).$$

Whenever $t_x < 0$, similarly, we obtain

$$s_x - t_x \leq \log \left(\frac{\lambda}{\int_{B_l} d(v)g(v)dv} \right).$$

Therefore, for any value that may take t_x , we obtain

$$s_x - t_x \leq \max \left\{ \log \left(\frac{\lambda}{\int_{\bar{B}_l} d(v)g(v)dv} \right), \log \left(\frac{\lambda}{\int_{B_l} d(v)g(v)dv} \right) \right\}. \quad (19)$$

Considering the statements (18) and (19), it follows that

$$\begin{aligned}s_x &\leq \frac{1}{2} \max \left\{ \log \left(\frac{\lambda}{\int_{\bar{A}_l} d(v)g(v)dv} \right), \log \left(\frac{\lambda}{\int_{A_l} d(v)g(v)dv} \right) \right\} \\ &\quad + \frac{1}{2} \max \left\{ \log \left(\frac{\lambda}{\int_{\bar{B}_l} d(v)g(v)dv} \right), \log \left(\frac{\lambda}{\int_{B_l} d(v)g(v)dv} \right) \right\} =: s_1. \quad (20)\end{aligned}$$

Taking into consideration the definition of the function V_x given in the statement (3), it is clear, for any $t \in \mathbb{R}$, that

$$V_d(t) := \frac{\int e^{tl(v)} l(v) (D(v)\mathbf{1}_{C_l}(v) + d(v)\mathbf{1}_{\bar{C}_l}(v))g(v)dv}{\int e^{tl(v)} D(v)g(v)dv} \leq V_x(t).$$

Consequently, considering the inverses of the functions V_d and V_x together with equations of the statement (17), since V_x is a nondecreasing function, it follows, whenever $\frac{\mu}{\lambda} > v_{0,d} := \inf_t V_d(t)$, that

$$t_1 := V_d^{-1} \left(\frac{\mu}{\lambda} \right) := \inf \left\{ s : V_d(s) \geq \frac{\mu}{\lambda} \right\} \geq \inf \left\{ s : V_x(s) \geq \frac{\mu}{\lambda} \right\} = t_x. \quad (21)$$

Observe now, for any $t \in \mathbb{R}$, that we have

$$V_D(t) := \frac{\int e^{tl(v)} l(v) (d(v) \mathbf{1}_{C_l}(v) + D(v) \mathbf{1}_{\bar{C}_l}(v)) g(v) dv}{\int e^{tl(v)} d(v) g(v) dv} \geq V_x(t).$$

Similarly as above, whenever $\frac{\mu}{\lambda} < v_{1,D} := \sup_t V_D(t)$, we have

$$t_0 := V_D^{-1} \left(\frac{\mu}{\lambda} \right) := \inf \left\{ s : V_D(s) \geq \frac{\mu}{\lambda} \right\} \leq \inf \left\{ s : V_x(s) \geq \frac{\mu}{\lambda} \right\} = t_x. \quad (22)$$

Moreover, it is obvious from the statement (17)(i) that

$$e^{s_x} \left(\int_{C_l} e^{t_0 l(v)} D(v) g(v) dv + \int_{\bar{C}_l} e^{t_1 l(v)} D(v) g(v) dv \right) \geq \lambda.$$

Therefore, for any $\lambda > 0$, we have

$$s_x \geq \log \left(\frac{\lambda}{\int_{C_l} e^{t_0 l(v)} D(v) g(v) dv + \int_{\bar{C}_l} e^{t_1 l(v)} D(v) g(v) dv} \right) =: s_0. \quad (23)$$

Consequently, making use of the statements (20), (21), (22) and (23), it is clear that there exist finite numbers S_0 and T_0 such that

$$\sup_{x \in \mathcal{C}} |\Gamma_x(\lambda, \mu) - \Gamma_x(\lambda_1, \mu_1)| \leq |\lambda - \lambda_1| |S_0| + |\mu - \mu_1| |T_0|.$$

This establishes that the family of functions $\{\Gamma_x(\lambda, \mu)\}_{x \in \mathcal{C}}$ is a equi-continuous and achieves the proof. \square

REFERENCES

1. BOSQ, D. (1998). Nonparametric Statistics for Stochastic Processes. Lecture Note in Statistics. Springer, New York.
2. DEMBO, A. & ZEITOUNI, O. (1998). *Large deviations techniques and applications*. Second edition, Springer-Verlag, New york.
3. DEUSCHEL, J.D. & STROOCK, D.W. (1989). *Large deviations*. Academic Press.
4. EZZAHRIQUI, M. AND OULD-SAÏD, E. (2008). Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *J. Nonparametric. Statist.*, **20**, 3–18.

5. FERRATY, F. AND VIEU, P. (2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. *C. R. Acad. Sci. Paris Sér. I Math.*, **330**, 139–142.
6. FERRATY, F. AND VIEU, P. (2004). Nonparametric models for functional data, with applications in regression, time series prediction and curve discrimination. The International Conference on Recent Trends and Directions in Nonparametric Statistics. *J. Nonparametric Statist.*, **16**, 111–125.
7. FERRATY, F., LAKSACI, A. AND VIEU, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat. Inference Stoch. Process.*, **9**, 47–76.
8. FERRATY, F. AND VIEU, P. (2006). *Nonparametric functional data analysis. Theory and practice*. Springer Series in Statistics. Springer, New York
9. FERRATY, F., MAS, A. AND VIEU, P. (2007). Nonparametric regression of functional data: inference and practical aspects. *Aust.N.Z.J.Statist.*, **49**, 267–286.
10. GASSER, T., HALL, P. AND PRESNELL, B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *J. Roy. Statist. Soc. Ser B*, **60**, 681–691.
11. LAÏB, N. AND LOUANI, D. (2010). Nonparametric Kernel Regression Estimation for Functional Stationary Ergodic Data: Asymptotic Properties. *J. Multivariate Analysis*, **101**, 2266–2281.
12. LOUANI, D. (1999). Some large deviations limit theorems in conditional nonparametric statistics. *Statistics*, **33**, 171–196.
13. LOUANI, D. AND OULD MAOULOUD, S. M. (2011). Some Functional Large Deviations Principles in Nonparametric Function Estimation. *J. Theoretical Probability*, **24**. In press.
14. NIKITIN, YA. (1995). *Asymptotic efficiency of non-parametric tests*. Cambridge University Press, Cambridge.
15. OULD MAOULOUD, S. M. (2008). Some uniform large deviation results in nonparametric function estimation. *J. Nonparametric Statist.*, **20**, 129–152.

16. VAN DER VAART, A. W. & WELLNER, J. A. (1996). Weak convergence and empirical processes. With applications to statistics. Springer Series in Statistics. Springer-Verlag, New York.